Artificial Intelligence & Autonomy

Opportunities and Challenges

Andrew Ilachinski

October 2017

CNA
ANALYSIS & SOLUTIONS

This document contains the best opinion of CNA at the time of issue.

It does not necessarily represent the opinion of the sponsor.

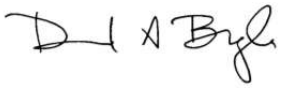
Distribution

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

Request additional copies of this document through inquiries@cna.org.

Photography Credits: Licensed image purchased from ShutterStock

Approved by: **October 2017**



Dr. David A. Broyles
Special Activities and Innovation Team
Operations Evaluation Group

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
10-2017	Final	

4. TITLE AND SUBTITLE
Artificial Intelligence & Autonomy Opportunities and Challenges

5a. CONTRACT NUMBER
N00014-16-D-5003

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
0605154N

6. AUTHOR(S)
Andrew Ilachinski

5d. PROJECT NUMBER
R0148

5e. TASK NUMBER
D180.00

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Center for Naval Analyses
3003 Washington Blvd
Arlington, VA 22201

8. PERFORMING ORGANIZATION REPORT NUMBER

DIS-2017-U-016388-Final

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DCNO, Resources and Assessments (N8)
Navy Department
Washington, D.C. 20350

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The military is on the cusp of a major technological revolution, in which warfare is conducted by unmanned and increasingly autonomous weapon systems. This exploratory study considers the state-of-the-art of artificial intelligence (AI), machine learning, and robot technologies, and their potential future military implications for autonomous (and semi-autonomous) weapon systems. Although no one can predict how AI will evolve or how it will affect the development of military autonomous systems, we can anticipate many of the conceptual, technical, and operational challenges that DOD will face as it increasingly turns to AI-based technologies. We identified four key gaps facing DOD as the military evolves toward an "autonomy era": (1) a mismatch of timescales between the pace of commercial innovation and DOD's acquisition process; (2) an underappreciation of the fundamental unpredictability of autonomous systems; (3) a lack of a universally agreed upon conceptual framework for autonomy; and (4) a disconnect between the design of autonomous systems and CONOPS development. We examine these gaps, provide a roadmap of opportunities and challenges, and identify areas of future studies.

15. SUBJECT TERMS
AI, unmanned systems, swarms, robots, modeling and simulation

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Knowledge Center/Tanya McCants
U	U	U	SAR	38	**19b. TELEPHONE NUMBER** *(include area code)* 703-824-2123

Abstract

The military is on the cusp of a major technological revolution, in which warfare is conducted by unmanned and increasingly autonomous weapon systems. This exploratory study considers the state-of-the-art of artificial intelligence (AI), machine-learning, and robot technologies, and their potential future military implications for autonomous (and semi-autonomous) weapon systems. Although no one can predict how AI will evolve or how it will affect the development of military autonomous systems, we can anticipate many of the conceptual, technical, and operational challenges that DOD will face as it increasingly turns to AI-based technologies. We identified four key gaps facing DOD as the military evolves toward an "autonomy era": (1) a mismatch of timescales between the pace of commercial innovation and DOD's acquisition process; (2) an underappreciation of the fundamental unpredictability of autonomous systems; (3) a lack of a universally agreed upon conceptual framework for autonomy; and (4) a disconnect between the design of autonomous systems and CONOPS development. We examine these gaps, provide a roadmap of opportunities and challenges, and identify areas of future studies.

Note: This paper is an extended version of the Executive Summary of a much longer recent CNA report on the opportunities and challenges of AI, robots, and swarms (henceforth referred to as the *CNA AI Report* in the main text): https://www.cna.org/CNA_files/PDF/DRM-2017-U-014796-Final.pdf.

This page intentionally left blank.

Contents

This page intentionally left blank.

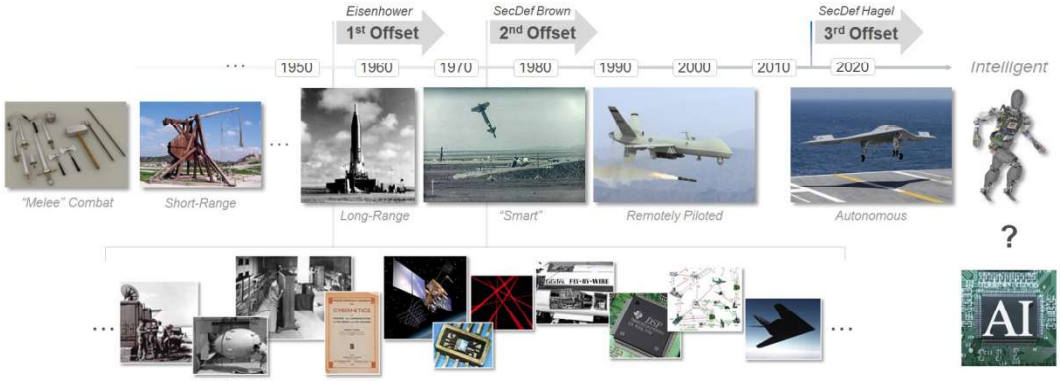
List of Figures

This page intentionally left blank.

Introduction

The history of warfare, in general, and the evolution of weapons, in particular, are both deeply entwined with developments in science and technology (see figure 1). Ever since a caveman first picked up a nearby rock to strike down an enemy, the use of new tools to safely project power from a distance has steadily evolved: from spears (which can be thrown as far as rocks but can inflict more damage); to bows and arrows (which significantly extended the offensive range and first appeared about 60,000 years ago); to catapults (which were introduced about 400 BC); to firearms (which were developed in China about 1200 and extended the range to hundreds of yards); to machine guns and modern artillery (which pushed the range still farther, to tens of miles), and air-dropped bombs (which were first deployed from balloons by Austria in 1849).

Figure 1. Schematic of the historical coupling between weapons and technology



In the modern era, in the U.S., an even deeper connection between the development of technology and weapons systems has been forged by national offset strategies. An *offset strategy* is a general set of peacetime policies designed to mitigate a perceived

tactical and/or strategic imbalance with one's main adversaries. For example, the First Offset, during President Eisenhower's administration in the 1950s at the start of the Cold War, was introduced to mitigate Soviet numerical and geographical advantages in Western Europe. Early digital computer technology and the burgeoning field of cybernetics led to the development of long-range intercontinental ballistic missiles (ICBMs) and enhanced air/missile defense networks.

During the Second Offset, introduced in the 1970s and 80s to mitigate the Soviet Union's newly established "peer status" with respect to nuclear weapon technology and delivery systems, strategic thinking turned to regaining a non-nuclear tactical advantage. Rapidly growing innovations in digital microelectronics and information technology led to the development of new intelligence, surveillance, and reconnaissance (ISR) platforms and battle management capabilities, precision-strike weapons, stealth aircraft, smart weapons and sensors, and the tactical exploitation of space (e.g., GPS).

The Third Offset—announced formally in November 2014 and centered on key investments in artificial intelligence (AI), human-machine collaboration, and autonomous unmanned systems—is designed to mitigate a *shrinking force structure* and *declining technological superiority.* The goal is not the acquisition of next-generation technologies, per se, but a combined re-evaluation of technological innovations and new concepts of operations. The major difference between this latest offset and its precursors is that whereas most First and Second Offset technologies were funded primarily by the Department of Defense (DOD), the key technology enablers are being developed almost exclusively in the commercial world.

A landmark event in Artificial Intelligence

A landmark AI-related event took place in March 2016: *AlphaGo*, a Go-playing AI developed by Google's *DeepMind*, defeated an 18-time world champion (Lee SeDol) in the game of Go.[3] For context, Go is a board game that was invented in China more than 2,500 years ago (making it the oldest board game in the world), and is played on a 19-by-19 grid onto which players alternate placing either white or black pieces (called "stones") in order to capture their opponent's territory (which is secured, and pieces "captured", when a board position is surrounded by a given color). *AlphaGo*'s victory over Lee SeDol is a landmark event because the number of possible moves in Go is so vast that, by almost any measure of complexity, it vastly exceeds that of

[3] C. Koch, "How the Computer Beat the Go Master," *Scientific American*, 19 March 2016.

chess![4] Indeed, prior to *AlphaGo*'s victory, most AI experts believed that no AI would defeat even a high-ranking human Go player for *another 15-20 years!*

Recall another landmark AI event that took place 20 years ago (in 1997), when IBM's chess-playing AI, *Deep Blue*, defeated the then-reigning human world champion in chess, Gary Kasparov.[5] It is instructive to underscore the differences between *Deep Blue*'s and *AlphaGo*'s respective achievements, and what each entails for the future development of AI in general. For example, *Deep Blue*'s core evaluation function—which is used to rank board positions—was handcrafted, albeit with many thousands of open parameter values, and later refined by chess grandmasters. The style of gameplay (which defeated Kasparov) was effectively "brute force," in which *Deep Blue* systematically applied its evaluation function to many alternative future states, searching seven or eight moves ahead for each player, at a rate of about 200 million position evaluations per second.[6]

The method used to build *AlphaGo* is very different from the one used for *Deep Blue*, and is a harbinger of the future of so-called "narrow AI" (i.e., non-sentient artificial intelligence, whose problem-solving capability is confined to one narrow task). Because of Go's complexity, a "brute force" approach is untenable—even a modest three-move "look ahead" requires close to 250 trillion board-position evaluations (compared to roughly two billion for chess).

Instead of codifying the expertise of human grandmasters, *AlphaGo* uses a pair of *neural networks* (NNs): one is trained to recognize good moves, and the other is trained to recognize good board positions. NNs are among the oldest, and most powerful, forms of "bottom up" AI methods.[7] Inspired by the way animal brains work, NNs consist of various layers of nodes (NNs with very many layers are referred to as *deep learning NNs,* or DLNNs). Data—in *AlphaGo*'s case, the position of the stones on a Go board—is presented to the first (input) layer, and then undergoes a series of mathematical transformations (the nature of which is determined by "weights" that connect individual nodes) as it flows through subsequent layers. NNs

[4] For example, using one standard metric, called *game tree complexity* (GTC)—which, roughly speaking, measures the number of positions a move-ranking algorithm would have to evaluate in order to determine the value of an initial position—the complexity of Go exceeds that of chess by almost 240 *orders of magnitude*. J. Burmeister, "The challenge of Go as a domain for AI research: a comparison between Go and chess," *Intelligent Information Systems*, 1995.

[5] Deep Blue beat G. Kasparov in 1997. See https://www.youtube.com/watch?v=NJarxpYyoFI.

[6] M. Campbell, "Knowledge discovery in Deep Blue," *Comm. of the ACM* 42, Nov. 1999.

[7] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," *Technical Report IDSIA-03-14*, arXiv:1404.7828 v4, 2014.

are "trained" to associate certain desired output states (such as a particular stone being moved from one position to another) with input patterns.

Whereas *Deep Blue* was essentially an expert system built using handcrafted rules, *AlphaGo* uses general machine-learning techniques to effectively *teach itself*. The two NNs described above were first trained on a database of about 30 million moves from games played by strong (though not anywhere near "world's best") human players. The second training phase consisted of *AlphaGo* playing itself several million times. In the match against SeDol, *AlphaGo* evaluated *thousands of times fewer positions* than *Deep Blue* did against its match against Kasparov. *AlphaGo*'s playing strength derives, instead, from a combination of self-learned evaluation of patterns (of stones) and policy for selecting only a relatively small set of board positions for "look ahead."

Notably, *no explicit programming was involved*. Indeed, the "rules" (which, for NNs, are best described as vast set of mathematical transformations, as there are no *rules* as such) by which *AlphaGo* selects its moves are unknown to any of its programmers. While this irreducible "unknowability" of how a trained NNs make their decisions is nothing new—it is a well-known characteristic of all neural-net based learning[8]—its appearance in future military AI-based weapon systems is all but guaranteed if similar machine-learning techniques are used, and raises the more ominous spectre of military autonomous systems sometimes behaving unpredictably.

It is also telling that both human victims (Kasparov, in his match against *Deep Blue,* and Lee SeDol in his match with *AlphaGo*), expressed genuine surprise at some point during their respective matches. While Kasparov was stunned by a human-like sacrifice of a pawn, which was later revealed to be a result of a programming error,[9] Lee SeDol was so stunned by the 37th move of the 2nd game—*"It's not a human move. I've never seen a human play this move. So beautiful."* [10]—that he had to leave the room for 15 minutes to recover his composure.

One key takeaway from *AlphaGo*'s victory over Lee SeDol is not that AI can now play the game of *Go* at a superhuman level, but that essentially the same learning method

[8] M. Hassoun, *Fundamentals of Artificial Neural Networks,* MIT Press, 2003.

[9] During an interview after the match, Kasparov was so "stunned" by *Deep Blue*'s seeming sacrifice of a pawn in the first game—"a wonderful and extremely human move"—that it altered the way he played in subsequent games, and, arguably, contributed to his eventual defeat. G. Kasparov, "The day that I sensed a new kind of intelligence," *Time*, 25 March 1996). However, 15 years later, one of *Deep Blue*'s designers suggested that the move was due to a programming bug: K. Finley, "Did a computer bug help Deep Blue beat Kasparov?," *Wired*, 28 Sep 2012.

[10] C. Metz, "The sadness and beauty of watching Google's AI play Go," *Wired*, 11 March 2016.

can be used to develop AIs that may equal or exceed humans at *anything easier than Go,* coupled with the realization that this includes a vast space of problems. Another key takeaway is subtle but no less important, and one that is not often included in the otherwise justifiably ebullient open press coverage of *AlphaGo*'s landmark achievement: the particular techniques used to develop *AlphaGo* (DLNNs, supervised learning, and reinforcement learning) represent but a small fraction of a much larger space of AI methods—including natural language processing, inferential reasoning, and knowledge representation—that, collectively, span a wide spectrum of maturity and "off the shelf" level of applicability. While *AlphaGo*'s victory remains notable for the reasons cited above, it is still limited to the space of well-defined "narrow AI" problems for which large training datasets are readily available. *AlphaGo*'s victory is thus a harbinger both of technologies to come and of challenges that await those who are developing them.

Other groundbreaking AI-related technologies

A notable number of groundbreaking AI-related technology announcements and/or demonstrations have taken place just since *AlphaGo*'s victory over Lee SeDol last year:

- AI learned—*on its own*—where to find the information it needs to accomplish a specific task.[11]
- AI predicted the immediate future (by generating a short video clip) by *examining a single photograph* (and can also predict the future from studying video frames).[12]
- AI automatically inferred the rules that govern the behavior of individual robots within a robotic swarm *simply by watching*.[13]
- AI learned to navigate the London Underground *by itself* (by consulting its own acquired memories and experiences, much like a human brain).[14]
- AI speech recognition reached human parity in conversational speech.[15]

[11] K. Narasimhan et al., "Improving Information Extraction by Acquiring External Evidence with Reinforcement Learning," presented at EMNLP 2016. https://arxiv.org/abs/1603.07954.

[12] C. Vondrick, H. Pirsiavash, and A. Torralba, "Generating Videos with Scene Dynamics," presented at the 29th Conference on Neural Information Processing Systems, Barcelona, Spain, 2016. http://web.mit.edu/vondrick/tinyvideo/paper.pdf.

[13] W. Li, M. Gauci, and R. Gross, "Turing learning: a metric-free approach to inferring behavior and its application to swarms," *Swarm Intelligence* 10, no. 3, September 2016. http://link.springer.com/article/10.1007%2Fs11721-016-0126-1.

[14] E. Gibney, "Google's AI reasons its way around the London Underground," *Nature*, Oct 2016.

- AI communication system *invented its own encryption scheme*, without being taught specific cryptographic algorithms (and without revealing to researchers how its method works).[16]
- AI translation algorithm invented its own "interlingua" language to more effectively translate between any two languages (*without being taught to do so by humans*).[17]
- AI system *interacted with its environment* (via virtual actuators) to learn and solve problems in the same way that a human child does.[18]
- AI-based medical diagnosis system at the Houston Methodist Research Institute in Texas achieved 99 percent accuracy in reviewing millions of mammograms (at a rate 30 times faster than humans).[19]
- AI poker-playing program defeated some of the world's best human poker players during a three-week-long tournament.[20]
- AI effectively "read minds" (of human test subjects looking at pictures of faces, via functional magnetic resonance images, or fMRI, of brain activity).[21]

These and other recent similar breakthroughs (e.g., IBM's *Watson's* defeat of the two highest ranked *Jeopardy!* players of all time in 2011),[22] are notable for several reasons. First, they collectively provide evidence that we, as a species, have already crossed over into an era in which seeing AI outperform humans—at least for specific

[15] X. Xiong et al., "Achieving Human Parity in Conversational Speech Recognition," *arXiv*, 2016. https://arxiv.org/abs/1610.05256.

[16] M. Abadi and D. Andersen, "Learning to Protect Communications with Adversarial Neural Cryptography," arXiv:1610.06918v1. https://arxiv.org/abs/1610.06918.

[17] Q. Le and M. Schuster, "A Neural Network for Machine Translation, at Production Scale," Google Research Blog, 27 Sep 2016. https://research.googleblog.com/2016/09/a-neural-network-for-machine.html.

[18] M. Denil, P. Agrawal, T. Kulkarni, et al., "Learning to perform physics experiments via deep reinforcement learning," under review as a conference paper to ICLR 2017. https://arxiv.org/pdf/1611.01843v1.pdf.

[19] T. Patel et al., "Correlating mammographic and pathologic findings in clinical decision support using NLP and data mining methods," *Cancer* 123, 1 Jan 2017.

[20] This event is noteworthy because success in poker requires a player to deal with imperfect information and an ability to "bluff" opponents. C. Metz, "Inside Libratus, the poker AI that out-bluffed the best humans," *Wired*, February 1, 2017.

[21] H. Lee and B. A. Kuhl, "Reconstructing Perceived and Retrieved Faces from Activity Patterns in Lateral Parietal Cortex," *Journal of Neuroscience* 36, no. 22, June 2016.

[22] S. Baker, *Final Jeopardy: Man vs. Machine and the Quest to Know Everything,* Houghton Mifflin Harcourt, 2011.

tasks—is *almost* routine (perhaps in the same way that landing on the moon was "almost" routine after the first few Apollo missions).[23] Second, they offer a glimpse of how *different* AI is from human intelligence, and how inaccessible its "thinking" is to outside probes. And third, they demonstrate the power of AI to *surprise* us (including AI system developers, who nowadays are closer in spirit to "data collectors" and "trainers" than to traditional programmers)—i.e., AI, at its core, is fundamentally *unpredictable*.

The breakthroughs listed above are also notable for a fourth reason—a more subtle one, but one that directly inspired this study. Namely, they portend a set of deep conceptual and technical challenges that the Department of Defense (DOD) must face, now and in the foreseeable future, as it embraces *AI*–, *robot*–, and *swarm*–related technologies to enhance (and weaponize) its fleet of unmanned systems with higher levels of autonomy. The subtlety lies in unraveling the true meaning of the deceptively "obvious" word, *autonomy;* indeed, as of this writing, there is no universally accepted definition. Unlike military innovations introduced during the Cold War era (stealth, GPS, precision guided munitions), the AI-based technology enablers of the 20XX-era are—and will likely continue to be—driven primarily by the commercial world.

Of course, the competition is also not standing still. On the asymmetric side of the warfare spectrum, terrorist groups can now easily deploy off-the-shelf drones with onboard sensors enhanced with freely available online image-recognition programs to provide real-time situational awareness. On the symmetric end of the spectrum, we need only note that, at the January 2017 meeting of the Association for the Advancement of Artificial Intelligence (AAAI)—for the first time—an equal number of peer-reviewed papers were accepted from China and the United States. The Chinese have also recently publically declared that their goal is to be the dominant AI power by 2030.[24]

[23] Unlike the Apollo program, however, AI is here to stay: *Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence*, Report of the 2015 Study Panel, Stanford University, Sep 2016.

[24] G. Webster et.al., "China's Plan to 'Lead' in AI: Purpose, Prospects, and Problems," *New America*, 1 Aug. 2017. https://www.newamerica.org/cybersecurity-initiative/blog/chinas-plan-lead-ai-purpose-prospects-and-problems/.

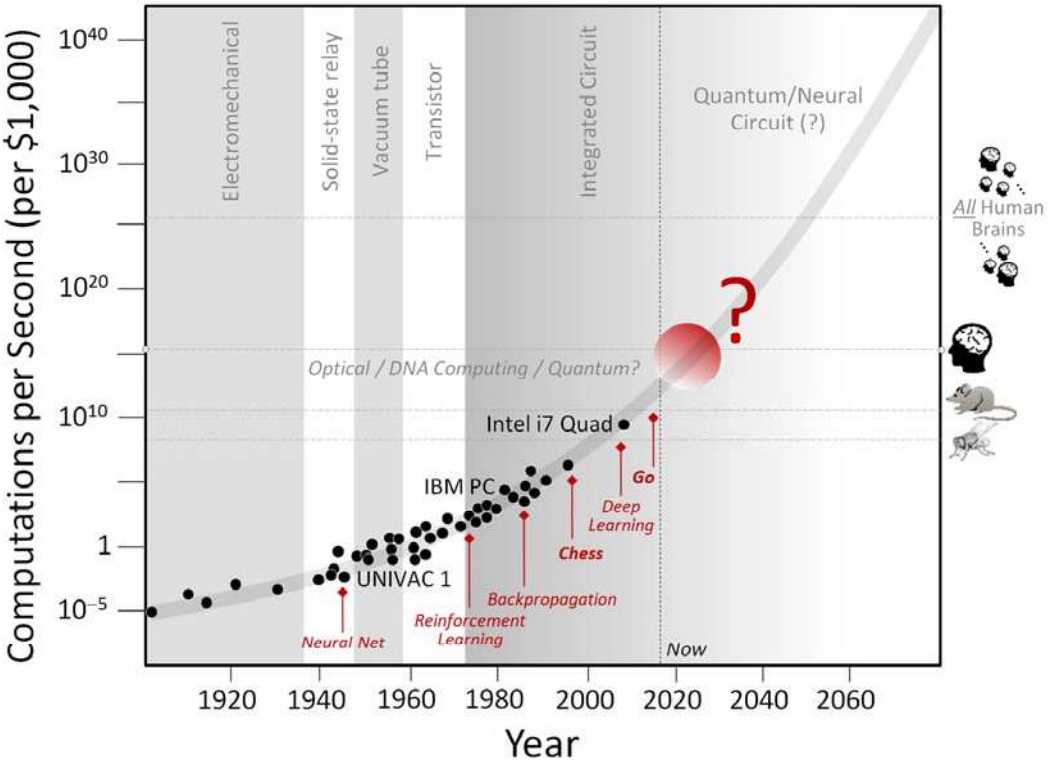
Accelerating technological change

The increasingly rapid pace of technological change, particularly in the fields of AI and robotics, is well known and documented.[25] For example, Ray Kurzweil (who directs Google's research on machine intelligence and natural language understanding), has argued that technology, like biology, is an evolutionary process whereby the information-processing tools and methods of prior generations are used to generate those of the next. As improvements accrue and evolve, the time between successive advancements in order and capability decreases exponentially.[26] Moreover, according to Kurzweil's "Law of Accelerating Returns," if a technology stalls or comes up against some form of barrier impeding further progress, a new technology will be invented to militate the presence of the barrier.

[25] *Timeline of Computer History*. http://www.computerhistory.org/timeline/ai-robotics/; J. Goodell, "Inside the Artificial Intelligence Revolution: A Special Report: Parts 1 & 2," *Rolling Stone Magazine*, Feb. and Mar. 2016.

[26] R. Kurweil, *The Singularity is Near*, Viking Press, 2005.

Figure 2. Accelerating growth of computing power



Ref: https://upload.wikimedia.org/wikipedia/commons/d/df/PPTExponentialGrowthof_Computing.jpg

Figure 2 shows, as a microcosm of a much larger space of general engineering and technology innovations,[29] a timeline of the accelerating growth of *computer power*, measured in terms of the number raw computations per second (CPS) that one can purchase for roughly $1,000.[30] Note that the CPS curve is logarithmic, meaning that each unit increment along the ordinate on the plot represents a jump of 10 times the prior value. Sequenced from the bottom to the top of the figure, the four horizontal dashed lines represent the approximate CPS values for an insect's brain, a mouse's brain, a human's brain, and all human brains on the planet, respectively.

[29] See, for example: T. Jackson, editor, *Engineering: An Illustrated History from Ancient Craft to Modern Technology*, Shelter Harbor Press, 2016.

[30] C. Moore and A. Mertens, *The Nature of Computation*, Oxford University Press, 2011.

The vertical dashed line, the bottom of which is buttressed against the label "Now," is centered on the year 2017 (i.e., this paper's release date). The alternating shades of gray, from left to right, represent overlapping technological epochs, and range from an era of *electromechanical devices*, to *solid-state relays*, *vacuum tubes*, *transistors*, and *integrated circuits*. In addition, six AI-related milestones are highlighted in red along the bottom of the CPS timeline curve (all are discussed later in this report): neural nets (introduced in 1943); reinforcement learning (a technique that is, in part, responsible for Google's *AlphaGo*'s recent defeat of Lee SeDol in Go), introduced in 1973; the backpropagation algorithm (which allows neural nets to "learn"), introduced in 1986; IBM Deep Blue's landmark victory over world champion Gary Kasparov in chess, developed in 1997; the "deep learning" algorithm (used by *AlphaGo* and other recent AI systems), published in 2006; and *AlphaGo*'s historic win over the reigning world champion human player in Go in 2016. The red disk that appears just to the right of center of the figure denotes the area of uncertainty in the expected continued growth in CPS in the coming decade.

One obvious takeaway from figure 2 is the observation that—as of this writing (Sep 2017)—the exponential CPS-vs-time curve is *about a decade away* from crossing the line that denotes the raw computational power of a human brain (the value of which is estimated to be between 10^{15} and 10^{16} CPS).[31] The precise value does not matter; nor does a "one human brain equivalent" of CPS represent a special barrier (such as the "speed of sound" for a jet) at which something magical happens. However, it does denote a computational threshold vastly beyond what our experience with computational technology has thus far prepared us for. The other takeaway is hidden behind what only *appears* to be a steady progression of capability from the "discovery" of neural nets (in the 1940s) to *AlphaGo*'s victory over Lee SeDol. The reality is that this progression was anything but steady, and was punctuated by at least two (decade-long) "dark periods" during which almost all research stopped, awaiting vital new "discoveries" (details in the *CNA AI Report*). Despite the recent rapid progress in narrow AI, it is not a given that all of the AI-based technologies required to develop fully autonomous weapon systems will advance at the same rate.

[31] N. Bostrom, "How long before superintelligence?" *Int. Jour. of Future Studies* 2, 1998.

Autonomous weapons

Autonomous weapons—colloquially speaking—have been used since World War II (e.g., the German *Wren* torpedo's passive acoustic homing seeker effectively made it the world's first autonomously guided munition).[32] Human-supervised automated defensive systems have existed for decades, and aerial drones were first used more than 20 years ago (i.e., the RQ-1 Predator was used as an intelligence, surveillance, and reconnaissance platform in former Yugoslavia).[33] But it was only after the September 11, 2001, terrorist attacks that the military's burgeoning interest in, and increasing reliance on, unmanned vehicles started in earnest. In just 10 years, DOD's inventory of unmanned aircraft grew from 163, in 2003, to nearly 11,000, in 2013 (and, in 2013, accounted for 40 percent of *all* aircraft).[34] And the United States is far from being alone in its interest in drones: by one recent tally, at least 30 countries have large military drones, and the *weaponized* drone club has recently grown to 11 nations, including the United States.[35]

DOD procured most of its medium-sized and larger unmanned aerial vehicles (UAVs), the MQ-1/8/9s and RQ-4s, for the counterinsurgency campaigns in Iraq and Afghanistan, where the airspace was largely uncontested. Now the United States is withdrawing from those campaigns and the military is shifting its strategic focus to less permissive operating environments (i.e., the Asia-Pacific region) and to adversaries with modern air defense systems. Thus, there is a growing emphasis on developing new, more autonomous, systems that are better equipped to survive in more contested airspaces.

Fundamentally, an autonomous system is a system that can independently compose and select among alternative courses of action to accomplish goals based on its knowledge and understanding of the world, of itself, and of the local, dynamic context. Unlike automated systems, autonomous systems must be able to respond to

[32] J. Campbell, *Naval Weapons of World War Two*, Naval Institute Press, 2002.

[33] P. Springer, *Military Robots and Drones: A Reference Handbook,* ABC-CLIO, 2013.

[34] *Unmanned Systems Integrated Roadmap: FY2013-2038,* U.S. Department of Defense, 2013.

[35] *World of Drones: Military*, International Security Data Site, New America Foundation. http://securitydata.newamerica.net/world-drones.html.

situations that are not pre-programmed or anticipated prior to their deployment. In short, autonomous systems are inherently, and irreducibly, *artificially intelligent robots.*[36]

To start, if and when autonomous systems, in the sense just described, finally arrive, they will offer a variety of obvious advantages to the warfighter. For example, they will eliminate the risk of injury and/or death to the human operator; offer freedom from human limits on workload, fatigue, and stress; and be able to assimilate high-volume data and make "decisions" based on time scales that far exceed human ability. If robotic swarms are added into the mix, entirely new mission spaces potentially open up as well—e.g., wide-area, long-persistence, surveillance; networked, adaptive electronic jamming; and coordinated attack. There are also numerous advantages to using swarms rather than individual robots, including: *efficiency* (if tasks can be decomposed and performed in parallel), *distributed action* (multiple simultaneous cooperative actions can be performed in different places at the same time), and *fault tolerance* (the failure of a single robot within a group does not necessarily imply that a given task cannot be accomplished).

Technical challenges

The design and development of autonomous systems entails significant conceptual and technical challenges, including:

- *"Devil is in the details" research hurdles:* Developers of autonomous systems must confront many of the same fundamental problems that the academic and commercial AI and robotic research communities have struggled for decades to "solve." To survive and successfully perform missions, autonomous systems must be able to sense, perceive, detect, identify, classify, plan for, decide on, and respond to a diverse set of threats in complex and uncertain environments. While aspects of all these "problems" have been solved to varying degrees, there is, as yet, no system that fully encompasses all of these features.

- *Complex and uncertain environments:* Autonomous systems must be able to operate in complex—possibly, a priori unknown—environments that possess a large number of potential states that cannot all be pre-specified or be exhaustively examined or tested. Systems must be able to assimilate, respond to, and adapt to dynamic conditions that were not considered during their

[36] A. Ilachinski, *AI, Robots, and Swarms*, CNA, DRM-2017-U-014796, January 2017. https://www.cna.org/CNA_files/PDF/DRM-2017-U-014796-Final.pdf.

design. This "scaling" problem—i.e., being able to design systems that are developed and tested in static and structured environments, and then have them perform as required in dynamic and unstructured environments—is highly nontrivial.

- *Emergent behavior:* For an autonomous system to be able to adapt to changing environmental conditions, it must have a built-in capacity to learn, and to do so without human supervision. It may be difficult to predict, and be able to account for *a priori* unanticipated, emergent behavior (a virtual certainty in sufficiently "complex" systems-of-systems dynamical systems).

- *Human-machine interactions/I:* The operational effectiveness of autonomous systems will depend on the dynamic interplay between the human operator and the machine(s) in a given environment, and on how the system responds, in real time, to changing operational objectives, in concert with the human's own adaptation to dynamic contexts. The innate unpredictability of the human component in human-machine collaborative performance only exacerbates the other challenges identified on this list.

- *Human-machine interactions/II:* The interface between human operators and autonomous systems will likely include a diverse space of tools that include visual, aural, and tactile components. In all cases, there is the challenge of translating human goals into computer instructions (e.g., "solving" a long-standing "AI problem" of natural language processing), as well as that of depicting the machine's "decision space" in a form that is understandable by the human operator (e.g., allowing the operator to answer the question, "Why did the system choose to take action X?").

- *Control:* As autonomous systems increase in complexity, we can expect a commensurate decrease in our ability to both predict and control such systems—i.e., the "spectre of complacency in complexity." As evidenced by the general nature of recent AI breakthroughs, there is a fundamental tradeoff: either the AI can achieve a given performance level (e.g., it can play the game Go as well as, or better than, a human), or humans can be able to understand how its performance is being achieved).

Apart from these innately technical challenges to developing autonomous systems, there is a set of concomitant acquisition challenges, the origin of which is a recent shift in DOD's innovation-related procurement practices. While the U.S. government has always played an important role in fostering AI research (e.g., ARPA, DARPA, NSF, ONR), most key innovations in AI, robotics, and autonomy are now being driven by

the *commercial sector*,[37] and at a pace that DOD's relatively plodding stove-piped acquisition process is ill equipped to accommodate: it takes 91 months (7.6 years), on average, from the start of an analysis of alternatives (AoA) study to initial operational capability (IOC).[38] Even information technology programs—under whose rubric most AI-derived acquisitions naturally fall—have averaged 81 months. By way of comparison, note that within roughly this same interval of time, the commercial AI research community has gone from just *experimenting* with (prototypes of dedicated hardware-assisted) deep learning techniques,[39] to beating the world champion in Go (along with achieving many other major breakthroughs).

Of course, DOD acquisition challenges, particularly for weapons systems that include a heavy coupling between hardware and software, have been known for decades.[40] However, despite numerous attempts by various stakeholders to address these challenges, the generic acquisition process (at least on the traditional institutional level) remains effectively unchanged. Whatever progress has been made in recent years derives more from *workarounds* instituted by DOD to facilitate "rapid acquisition" of systems,[41] than from wholesale changes applied to stove-piped processes of the acquisition process itself. Some recent progress has been made—e.g., the 2009/2011 National Defense Authorization Acts (NDAA/Sec 804), mandated a new IT acquisition process, which, in turn, led to multiple Defense Science Board (DSB) Task Force (TF) studies of the acquisition process. Yet, a notable absence in any of these DSB/TF studies is any explicit mention of autonomy.

[37] The development of most of the UAVs used in Iraq and Afghanistan was driven not by DOD requirements, but rather by commercial research and development. "Microsoft, Google, Facebook and more are investing in artificial intelligence: What is their plan and who are the other key players?" *TechWorld*, Sep. 29, 2016.

[38] *Policies and Procedures for the Acquisition of Information Technology*, Department of Defense, Defense Science Board, Task Force Report, Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, Mar. 2009.

[39] The first graphics-processor-based unsupervised deep-learning techniques were introduced in 2009. R. Raina, A. Madhavan, and A. Ng, "Large-scale deep unsupervised learning using graphics processors," *Proceedings of the 26th Annual International Conference on Machine Learning*, ACM, 2009.

[40] J. Merritt and P. Sprey, "Negative marginal returns in weapons acquisition," *in American Defense Policy*, Third Edition, edited by R. Head and E. Roppe, John Hopkins Univ. Press, 1973.

[41] Examples include the U.S. Air Force Rapid Capabilities Office, the U.S. Army's Asymmetric Warfare Group and Rapid Capabilities Office, DOD's Strategic Capabilities Office, and, most recently, SecDef Ashton Carter's Defense Innovation Unit Experimental (DIUx). B. Fitzgerald, A. Sander, J. Parziale, *Future Foundry: A New Strategic Approach to Military-Technical Advantage*, Center for a New American Security, 2016.

Complicating the issue still further is a basic dichotomy between DOD's existing directive on autonomy (DOD Directive 3000.09, issued Nov 2012) and current practices in Test and Evaluation (T&E) and Verification and Validation (V&V). Specifically, Directive 3000.09 requires that weapons systems (italics added by author of this report):[42]

- Go through rigorous hardware and software T&E/V&V, "including analysis of *unanticipated emergent behavior* resulting from the effects of complex operational environments on autonomous or semiautonomous systems."
- "Function as anticipated in realistic operational environments against *adaptive adversaries.*"
- "Are sufficiently robust to minimize failures that could lead to *unintended engagements.*"

Directive 3000.09 further requires that T&E/V&V must "assess system performance, capability, reliability, effectiveness, and suitability under realistic conditions, including possible adversary actions, consistent with the *potential consequences of an unintended engagement or loss of control of the system.*"

Yet, existing T&E/V&V practices do not make accommodations for any of the italicized parts of these quoted requirements. Among the many reasons why autonomous systems are particularly difficult to test and validate are: (1) *complexity of the state-space* (it is impossible to conduct an exhaustive search of the vast space of possible system "states" for autonomous systems); (2) *complexity of the physical environment* (the behavior of an autonomous system cannot be specified—much less tested and certified—in situ, but must be tested in concert with interaction with a dynamic environment, rendering the space of system inputs/outputs and environmental variables combinatorically intractable); (3) *unpredictability* (to the extent that autonomous systems are inherently complex adaptive systems, novel or unexpected behavior can be expected to arise naturally and unpredictably in certain dynamic situations; existing T&E/V&V practices do not have the requisite fidelity to deal with emergent behavior); and (4) *human operator trust in the machine* (existing T&E/VV&A practice is limited to testing systems in closed, scripted environments, since "trust" is not an innate trait of a system).

Trust also entails grappling with the issue of *experience* and/or *learning:* to be more effective, autonomous systems may be endowed with the ability to accrue information and learn from experience. But such a capability cannot be certified monolithically, during one "check the box" period of time. Rather, it requires periodic

[42] Enclosures 2 and 3 of DoD Directive 3000.09 (*Autonomy in Weapon Systems*, Nov 2012) address T&E and V&V issues, and generally review guidelines, respectively.

retesting and recertification, the periodicity of which is necessarily a function of the system's history and mission experience. Existing T&E/V&V practices are wholly inadequate to address these issues.

Defining *autonomy*

"Autonomy" applies to a vastly greater range of processes than those that pertain to unmanned vehicles—as physical entities—alone, including the myriad factors needed to describe human-machine interactions. It represents a range of *context-dependent capabilities,* which may appear at different scales and in varying degrees of sophistication, that collectively enable the coupled human-machine system to perform specific tasks. Autonomy—by itself—does not reductively "fix" any existing problems; rather, it redefines, extends, and potentially opens up entirely new mission spaces. And its value can only be assessed in the context of specific mission requirements, the operating environment, and its coupling with human operators.

A major impediment to the development of autonomous weapon systems is the current lack of a common language by which AI, robot, and other technology experts, systems developers, and program managers can communicate (in a manner consistent with autonomy's multi-dimensional, context-dependent nature). There is not an even a single definition of the word *autonomy*, much less a universally agreed-upon taxonomy that might be used as basis for forming a common language. Some taxonomies emphasize the details related to a system's output functions (i.e., to its decision capability), while others focus on making detailed distinctions between input functions, such as how a system acquires information and how it formulates options. And, while sliding scales have been used to delineate between levels of "human control" that a given system might require (e.g., the "autonomy" of a system may be ranked from, say, 0, meaning that it is under complete control, to 10, meaning it is fully autonomous, albeit, typically, without the term *fully* being well defined), the practical utility of these kinds of taxonomies is limited because they ignore critically important contextual factors. For this reason, a recent U.S. Defense Science Board report recommended doing away with defining levels of autonomy altogether and replacing such taxonomies with a comprehensive conceptual framework. However, to date, despite a handful of ongoing attempts, no useable framework yet exists.

Ethical concerns

The emerging use of autonomous weapons—and the spectre (if not yet the reality) of *lethal* autonomous weapon systems (LAWS), that can select and engage targets on

their own[43]—raises a host of ethical and moral questions. For example, "Will soldiers be willing to go to battle alongside robots?" "Will robots be able to distinguish between military and civilian targets, and be able to use force proportionately?" "Will an AI be able to recognize enemy signs of surrender?" "Who will be responsible for an unjustified robotic kill?" and "How does one codify an innately subjective body of ethical standards and practices?"

Such questions have led to several international movements against "killer robots."[44] For example, in July 2015, over 1,000 robotics and artificial intelligence researchers signed an open letter calling for a ban on offensive autonomous weapons (with 20K+ signatories as of Dec 2016).[45] And, at the most recent United Nations Convention on Conventional Weapons, the 123 participating nations voted to convene a group of government experts to meet (during two sessions) in 2017 to formally address the LAWS issue, which could potentially lead to an international ban.[46]

While the outcome of these upcoming meetings is uncertain, it is clear is that the political, cultural, and basic human-rights dimensions of this issue are only beginning to be explored. An analysis of the *operational* impact that any limitations on (or an outright ban of) the use of offensive autonomous weapons may entail for U.S. military forces obviously deserves attention.

Transitioning to new autonomy-enabled mission areas

Figure 3 illustrates, schematically, the key steps involved in extending the existing unmanned systems mission space (e.g., reconnaissance, route clearance, and search

[43] Although there are a number of weapon systems in use today that depend on varying degrees of human supervision, there are none that are fully autonomous (with the only possible exception being the Israel Defense Forces *Harpy*, a "fire-and-forget" loitering munition designed to detect, attack and destroy radars). Autonomy policy for U.S. weapon systems is spelled out in DoD Directive 3000.09, which expressly prohibits use of lethal *fully* autonomous weapons, which it defines as weapon systems that, once activated, may select and engage targets without further intervention by a human. DoD Directive 3000.09, "Autonomy in Weapon Systems," Nov 2012. http://www.dtic.mil/whs/directives/corres/pdf/ 300009p.pdf.

[44] M. Wareham and S. Goose, "The Growing International Movement Against Killer Robots," *Harvard International Review*, 5 Jan 2017.

[45] http://futureoflife.org/open-letter-autonomous-weapons/.

[46] Final Document of the Fifth Review Conference, CCW, Dec 2016. http://www.reaching critical will.org/disarmament-fora/ccw/2016/revcon.

and rescue) to one that more fully embraces all that autonomy potentially offers (e.g., self-organized, and self-healing, adaptive swarms). Leaving aside details of the pipeline to the main text, the key (mutually entwined) steps include, starting from bottom of the figure and working our way to the top:

- *Step 1:* Conducting basic AI research across multiple domains (the green-to-red overlay emphasizing that research in different AI areas—e.g., deep learning, image recognition, and robotic swarms—necessarily proceeds at different rates and exists, at any one time, at different levels of maturation).

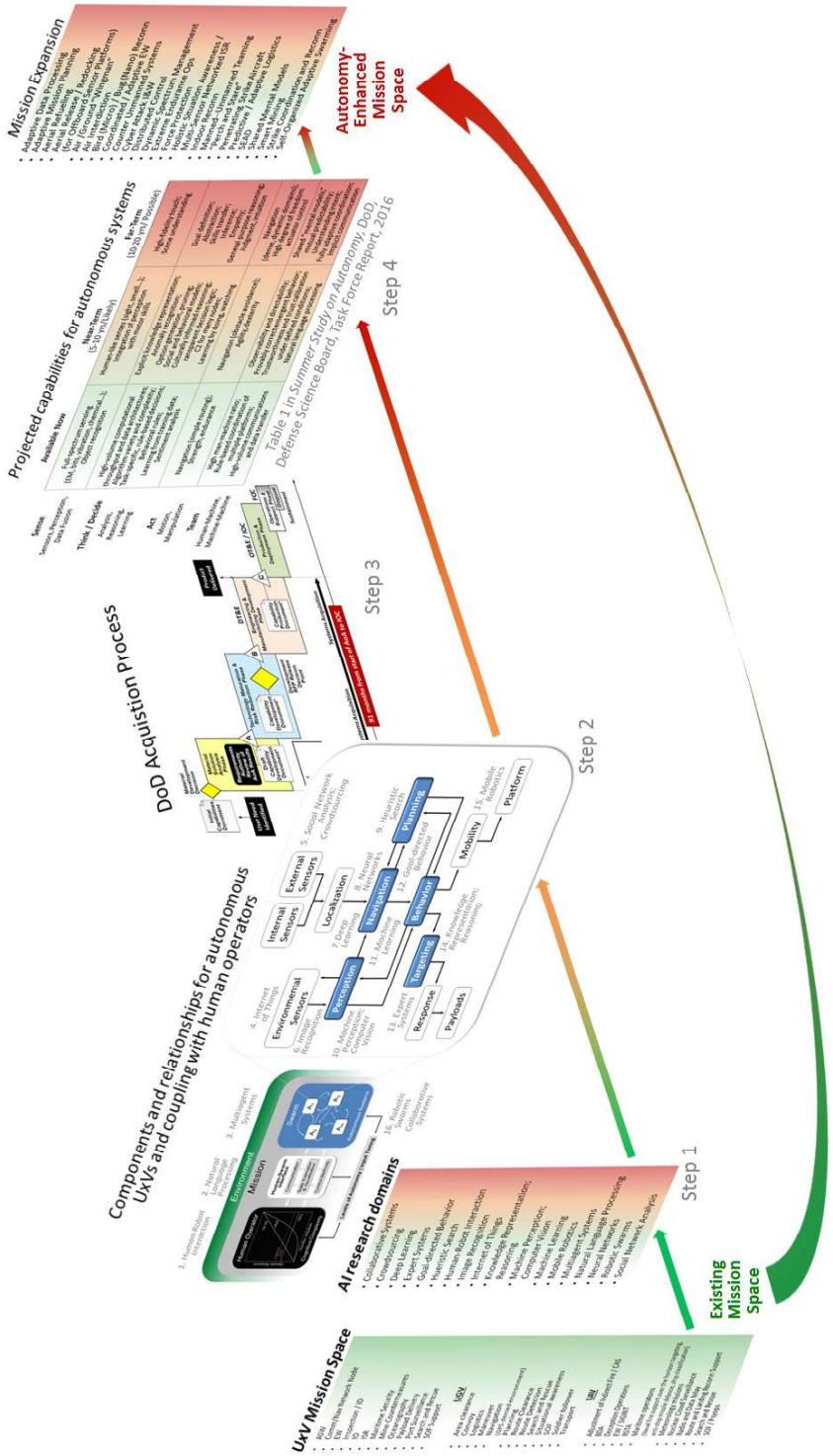
- *Step 2:* Understanding how individual AI research domains feed into the myriad components that make up autonomous systems, including their coupling with human operators (which further involves the understanding of how human-machine collaborative systems function in specific mission environments).

- *Step 3:* Moving design, development, testing, and accreditation through the DOD acquisition process (and accommodating autonomy's unique set of technical challenges while doing so).

- *Step 4:* Interpreting and projecting the requisite levels of maturity of system capabilities that autonomous systems must possess for specific missions. The autonomous systems that are shown in figure 2 are characterized as functions of four broad categories of AI (i.e., *sensing*, *thinking*, *acting*, and *teaming*). Their projected capabilities are indicated as follows: shades of green indicate capabilities that are available now; shades of orange denote near-term capabilities; and increasingly darker shades of red indicate the far-term regime. This table is taken from the DOD's Defense Science Board's most recent study on autonomy,[47] but is intended mostly as a notional place-holder for the kinds of conceptual, technical, and analytical considerations that must be taken into account as the raw capabilities of the autonomous systems that come out of the acquisition process are transformed into new and operationally meaningful missions and missions areas.

[47] Table 1 in *Summer Study on Autonomy*, Department of Defense, Defense Science Board, Task Force Report, Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, June 2016. https://www.hsdl.org/?view&did=79464.

Figure 3. Key steps in transitioning to new autonomy-enabled mission areas



Gestalt of main findings

The military is on the cusp of a major technological revolution as it enters the *Robotic Age,*[49] in which warfare is conducted by unmanned and increasingly autonomous weapon systems, operating across all domains (air, sea, undersea, land, space, and cyber), and across the full spectrum of military operations. The question is not *whether* the future of warfare will be filled with autonomous, AI-driven robots, but *when* and in what *form*. However, unlike the last "sea change" during the Cold War (i.e., the so-called "2nd Offset"),[50] when advanced technologies such as precision-strike weapons, stealth aircraft, smart weapons and sensors, and GPS were developed primarily by DOD-sponsored research and development programs, a successful transition into the Robotic Age (spurred on by DOD's recent "Third Offset Strategy" innovation initiative)[51] will depend critically on how well DOD is able to embrace technologies and innovations that are now being developed mostly in the commercial world. And, while the human warfighter is not going away anytime soon, if ever (even as the depth and breadth of autonomy steadily expand), human operators will not suddenly lose control of existing unmanned systems.

A telltale sign that DOD has made a "no looking back" cross-over into the Robotic Age will be when human operators can no longer fully understand, or *predict*, how autonomous systems behave—i.e., when, for the first time, a human operator is as stunned by some weapon system's action as 18-time world Go champion Lee SeDol was by a single move of the AI that defeated him.

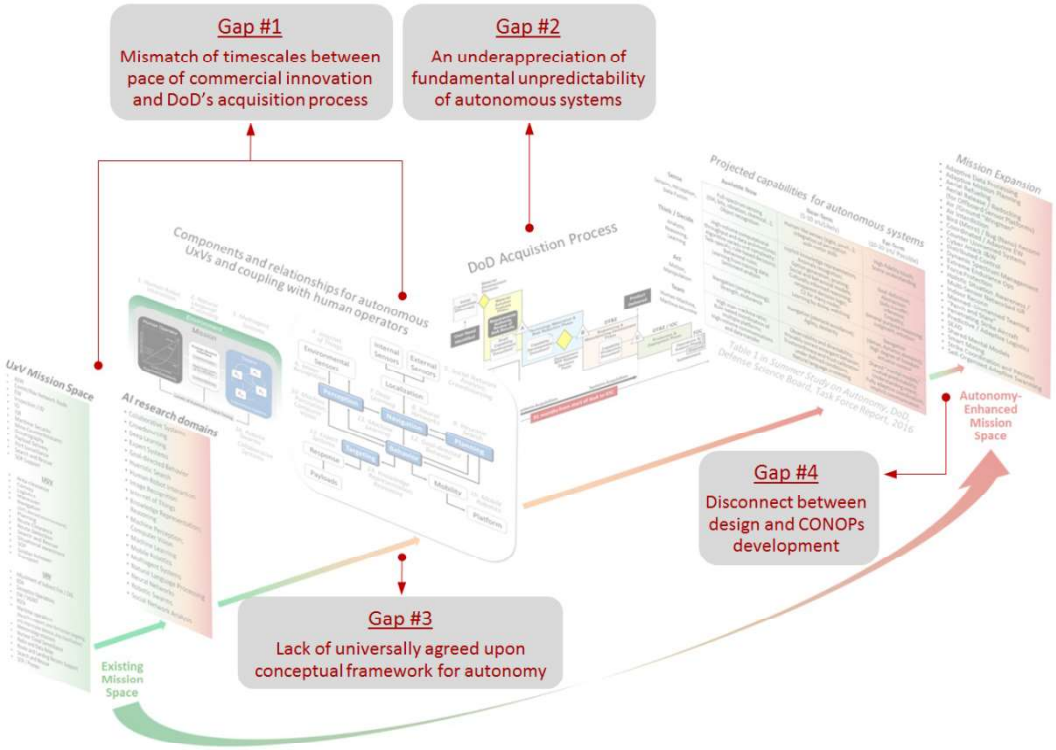
In preparation for DOD's cross-over into the Robotic Age, whenever it arrives, this study has identified four key technical gaps in developing AI-based autonomous systems, wherein opportunities for future analytical studies naturally arise (see figure 4).

[49] Robert O. Work and Shawn Brimley, *20YY: Preparing for War in the Robotic Age*, Center for a New American Security, Jan 2014.

[50] J. McGrath, "Twenty-First Century Information Warfare and the Third Offset Strategy," *Joint Forces Quarterly*, National Defense University, Issue 82, 3rd Quarter 2016.

[51] C. Hagel, Transcript of Keynote speech delivered at Reagan National Defense Forum Keynote, Ronald Reagan Presidential Library, Simi Valley, CA, Nov. 15, 2014.

Figure 4. Key gaps in transitioning to new autonomy-enabled mission areas



These gaps are:

- *Gap 1*: A fundamental mismatch—even *dissonance*—between the accelerating pace (and manner of development and evolution) of technology innovation in commercial and academic research communities, and the timescales and assumptions underlying DOD's existing acquisition process.

- *Gap 2:* An underappreciation of the unpredictable nature of autonomous systems, particularly when operating in dynamic environment, and in concert with other autonomous systems. Existing T&E/V&V practices accommodate neither the basic properties of autonomous systems, as expected by AI and indicated by decades of deep fundamental research into the behavior of complex adaptive systems, nor the requirements they must meet, as weapon systems (as spelled out by DOD Directive 3000.09).

- *Gap 3:* A lack of a universally agreed-upon conceptual framework for autonomy that can be used both to anchor theoretical discussions and to serve as a frame-of-reference for understanding how theory, design, implementation,

testing, and operations are all interrelated. A similar deficiency exists for understanding the role that trust plays in shaping a human operator's interaction with an autonomous system. The Defense Science Board's most recent study on autonomy[52] warns that "inappropriate calibration" of trust during "design, development, or operations will lead to misapplication" of autonomous systems, but offers only a tepid definition of trust, and little guidance on how to apply it.*Gap 4*: DOD's current acquisition process does not allow for a timely introduction of "mission-ready" AI/autonomy, and there is a general disconnect between system design and the development of concepts of operations (CONOPS). Unmanned systems are typically integrated into operations from a *manned*-centric CONOPS point of view, which is unnecessarily self-limiting by implicitly respecting human performance constraints.

[52] *Summer Study on Autonomy*, Department of Defense, Defense Science Board, Task Force Report, Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, June 2016. https://www.hsdl.org/?view&did=79464.

Recommended studies

While not even AI experts can predict how AI will evolve in even the near–term future (much less project its possible course over 10 or more years,[53] or predict AI's impact on the development of military autonomous systems), it is still possible to anticipate many of the key conceptual, technical, and operational challenges that DOD will face in the coming years as it increasingly turns to and more deeply embraces AI-based technologies, and fully enters the "Robotic Age." From an operational analysis standpoint, these challenges can also be used to help shape future studies:

- **Recommendation 1:** ***Help establish dialog between commercial research and development and DOD.***

Institutions specializing in operational analysis are well suited to act as "go betweens" linking the academic and commercial research communities with military culture / operational needs. Assuming that Secretary of Defense Ashton Carter's Defense Innovation Unit-Experimental (DIUx) program survives into the next administration,[54] operationally informed and technically knowledgeable analysts can help stakeholders better "understand" each other. Cross-fertilization with the Naval Postgraduate School (NPS) may also pay dividends.[55]

- **Recommendation 2:** ***Develop an operationally meaningful conceptual framework for autonomy.***

[53] S. Armstrong, K. Sotala, and S. Éigeartaigh, "The errors, insights and lessons of famous AI predictions – and what they mean for the future," *Journal of Experimental & Theoretical Artificial Intelligence* 26, no. 3, 2014; D. Fagella, "Artificial Intelligence Risk – What Researchers Think is Worth Worrying About," *Tech Emergence,* 20 March 2016. http://techemergence.com/artificial-intelligence-risk/. For the most recent survey of expert opinion see: V. Muller and N. Bostrom, "Future Progress in Artificial Intelligence: A Survey of Expert Opinion," in *Fundamental Issues of Artificial Intelligence*, edited by V. Muller, Springer-Verlag, 2016.

[54] DIUx has been established to help facilitate the discovery and development of capabilities and technologies outside DOD's normal acquisition pipeline. https://www.diux.mil/.

[55] For example: NPS's Consortium for Robotics and Unmanned Systems Education and Research (CRUSER: https://my.nps.edu/web/cruser), and *Autonomous Systems Track* (http://my.nps.edu/web/ast).

Despite rapid advances and DOD's growing interest in artificial intelligence (AI), robotics, and swarm technologies, a major current gap in the development of autonomous systems is a lack of a comprehensive universally agreed upon conceptual framework for autonomy that can be used both to anchor theoretical discussions and to serve as a frame-of-reference for understanding how theory, design, implementation, testing, and operations are all interrelated. In order to be useful, such a framework must be able to both objectively (or, as objectively as possible) distill and convolve all key elements of autonomy, and flexible and deep enough to anticipate and accommodate the development of future systems. To appreciate how technically difficult a task it is to find an appropriate set of metrics to describe both *what* an autonomous system is and how *well* it is performing, it is enough to recognize that the closer systems come to achieving full autonomy, the more closely aligned will any description of their behavior be to that of *describing the behavior of humans*. Therefore one ought not be surprised to learn that the autonomy-related research literature is replete with just about every combination of factors that may be used to categorize human, machine, and human-machine (hybrid) behaviors.

Of the frameworks that have been proposed in the literature, most are either far too shallow (such as the skeleton of an idea proposed by DOD's Defense Science Board's 2012 report on autonomy[56]) or too general (and/or subsequently stalled) efforts, such as the National Institute of Standards and Technology's (NIST's) ALFUS (Autonomy Levels for Unmanned Systems) framework.[57] No framework proposed to date has been developed specifically with DOD's unique requirements (vis-à-vis autonomous weapon systems) in mind. Recommendation #2 is to develop an operationally meaningful suite of human-machine-centric mission metrics and conceptual framework for AI-based autonomy, and show how it can be used to help support various components of DOD's acquisition process.

- **Recommendation 3:** ***Develop measures of effectiveness (MOEs) and measures of performance (MOP) for autonomous systems.***

Develop a methodology by which the effectiveness of autonomous systems can be measured at all levels (e.g., developers, program managers, decision-makers, and warfighters) and across all required functions, missions, and tasks (e.g., coordination, mission tasking, training, survivability, situation awareness, and workload).

[56] *The Role of Autonomy in DoD Systems*, DoD Defense Science Board, Task Force Report, Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, July 2012.

[57] https://www.nist.gov/el/intelligent-systems-division-73500/cognition-and-collaboration-systems/autonomy-levels-unmanned.

- **Recommendation 4:** ***Use nontraditional modeling and simulation (M&S) techniques to help mitigate AI/autonomy-related dimensions of uncertainty.***

As DOD moves into the Robotic Age, M&S is moving away from "simulations as distillations" of real systems (for which M&S has traditionally been used to develop models in order to gain insights into the *real* system), to "simulation-based rules and algorithms as descriptions" of real (i.e., engineered) robots and behaviors. It is here, on the cusp between exploring behaviors and prescribing rules that generate them (e.g., engineering *desired* swarm behaviors), that M&S can help mitigate some of the challenges and uncertainties of developing autonomous systems and robotic swarms. For example, while "swarm engineering" methods exist to facilitate the unique design requirements of robotic swarms, no general method exists that maps individual rules to (desired) group behavior.[58]

Multi-agent based modeling techniques[59] are particularly well suited for developing these rules, and, more generally, for studying the kinds of self-organized emergent behaviors expected to arise in coupled autonomous systems (e.g., "How sensitive is an autonomous system's behavior to changes in its physical environment?" "What new command and control architectures will be needed for robotic swarms?" and "How will the control and behavior of a swarm scale with its size and mission complexity?").

- **Recommendation 5:** ***Apply wargaming techniques to help develop new CONOPS.***

Wargaming can be used to help identify and develop new CONOPS, apply lessons-learned from the experience of using deployed systems, explore options to counter uses of autonomy by potential adversaries, and assist in training (e.g., by exploring trust issues in human-machine collaboration). Wargames can also stimulate and nurture a more unified approach to understanding autonomous system performance and behavior, provided that they are conducted with support and participation from across all military services and domains.

[58] I. Navarro and F. Matia, "An Introduction to Swarm Robotics," *International Scholarly Research Notes*, Vol. 2013, 2013. https://www.hindawi.com/ journals/isrn/2013/608164/.

[59] A. Ilachinski, *Artificial War: Multiagent-Based Simulation of Combat*, World Scientific, 2004. See also: A. Ilachinski, "Modelling insurgent and terrorist networks as self-organized complex adaptive systems," *International Journal of Parallel, Emergent and Distributed Systems* 27, 2012; A. Ilachinski, *AOEWSim: An Agent Based Model for Simulation Interactions Between Off-Board EW Systems and Anti-Ship Missiles*, CNA, DWP-2013-U-004757, 2013; A. Ilachinski and M. Shepko, *FAC/FIAC Simulation (FFSim): User's Guide*, CNA, Annotated Briefing, 2015.

- **Recommendation 6:** ***Develop new T&E/V&V standards and practices appropriate for the unique challenges of accrediting autonomous systems.***

For example, help ameliorate basic gaps in testing in terms of accommodating complexity, uncertainty, and subjective decision environments, by appealing to and exploiting lessons learned from the development and accreditation practices established by the complex system theory and multiagent-based modeling research communities.

- **Recommendation 7:** ***Assess the data requirements for developing machine-learning-based autonomy.***

As DOD inexorably moves toward developing and deploying autonomous weapon systems that will increasingly rely on AI technologies, it also faces a host of technical challenges. Not the least of these challenges is the availability—or, more precisely, the potential lack of—data required for AI systems to learn from. For example, the well-publicized defeat in 2016 of an 18-time world champion in Go by Google's *AlphaGo* Go-playing program derives, in large part, from a very large dataset of both human and machine-generated games (in the latter case, numbering in the millions) from which *AlphaGo* learned. Similarly, the better-than-human-level performance that image recognition and classification programs have recently achieved owes itself to the ready availability of massive training datasets (e.g., the image classifiers that competed in the 2014 Large Scale Visual Recognition Challenge (LSVRC), all trained on a set of images distributed among 1,000 categories and 1.2 million images; and training that required significant human effort to provide a large enough sample space of "correct" labels). Judging by recent government-sponsored reports on AI and autonomy, an analysis of issues having to do with the amount and kind of data necessary to train military AI systems has thus far garnered little attention. Recommendation #7 is to explore the unique DOD-driven data requirements for developing machine-learning-based autonomy.

- **Recommendation 8:** ***Explore basic human-machine collaboration and interaction issues.***

As autonomy increases, human operators will be concerned less with the manual control of a vehicle, and more with controlling swarms and directing the overall mission: "What are the operator's informational needs (and workload limitations) for controlling multiple autonomous vehicles?" "How do humans keep pace with an accelerating pace of autonomy-driven operations?" "What kinds of command-and-control relationships are best for human-machine collaboration?" "How are human and autonomous-system decision-making practices optimally integrated?" and "What data practices are key to developing shared situation awareness?"

- **Recommendation 9:** ***Explore the challenges of force-integration of increasingly autonomous systems.***

Essentially all force-integration issues are, as yet, undetermined. They must consider not just "low hanging fruit" extensions of existing CONOPS, in which the human component is simply replaced with unmanned systems and "operational value" of human performance is scaled to accommodate "better" performance (e.g., endurance, survivability), but brainstorm heretofore nonexistent tactics, operations, and missions that fully embrace existing and anticipated future autonomous capabilities. What is the tradeoff between large numbers of simple, low-cost (i.e., "disposable") vehicles and small numbers of complex (multi-functional) ones?

The operationalization of robotic swarms, in particular, represents a heretofore largely untapped dimension of the mission space, and will require the development of new CONOPS. The swarm may be used as a radically new form of precision coordinated "en masse" guided munition; as a self-healing area surveillance network (which includes collecting and assimilating data on an adversary's Internet-of-Things (IoT);[60] or as an adaptive distributed electronic jammer.

- **Recommendation 10:** ***Explore the cyber implications of autonomous systems.***

Explore what new features increased AI-driven autonomy brings to the general risk assessment of increasingly autonomous unmanned systems. On one hand, autonomy may potentially reduce a force's overall vulnerability to jamming or cyber hacking. For example, communications loss over a jammed data link may be compensated for by the ability of autonomous vehicles to continue performing their mission. On the other hand, autonomy itself may also be *more*, not less, vulnerable to a cyber intrusion. For example, an adversary may gain "control," or otherwise deliberately "perturb" the behavior of an autonomous system; it may also be more difficult to detect embedded malware. In the latter context, consider some future variants of incidents such as the Iranian capture of an RQ-170 *Sentinel* in 2011,[61] and the "keylogging" virus that infected the UAV-control-computers at Creech Air Force Base in Nevada.[62]

- **Recommendation 11:** ***Explore operational implications of ethical concerns over the use of lethal autonomous weapons.***

Analyze issues of accountability, legality, and liability in arguments put forth by various "Ban LAWS" movements. Examine the possible constraints on missions (along

[60] G. Seffers, "Defense Department Awakens to Internet of Things," *Signal*, 1 Jan 2015. http://www.afcea.org/content/?q=defense-department-awakens-internet-things.

[61] The Iranian government announced that the RQ-170 had been captured by its cyber warfare unit: "Iran shows film of captured US drone," BBC News, 8 Dec 2011. http://www.bbc.com/news/ world-middle-east-16098562.

[62] N. Shachtman, "Exclusive: Computer virus hits U.S. drone fleet," *Wired*, 7 Oct 2011.

with other associated impediments to the design and development of autonomous systems) that may result from an international ban (or set of limits) imposed on the development or deployment of LAWS, such as might come out of the government experts' negotiations, sponsored by the United Nations, that is scheduled to take place sometime in 2017.

Finally, lest the reader be left with the false impression that developments on the AI front impact unmanned autonomous systems alone:

- **Recommendation 12:** ***Examine the full spectrum of possible applications of expected near- to mid-term AI advances (for all military services).***

AI is a vast field, consisting of numerous overlapping methods and technologies in varying degrees of maturity and capability. Its continued development as a whole is certain to impact a much wider range of mission areas and operational needs and capabilities than those associated with autonomous weapon systems alone. Such areas include: basic pattern recognition for enhanced situational awareness; INTEL processing; exploitation and dissemination (PED) and prediction (e.g., the DOD's recently stood-up project Algorithmic Warfare Cross-Functional Team, designed to accelerate integration of big data and machine learning); real-time tactical decision aids (for deployment in both physical and virtual spaces); adaptive command and control; networked communicating Internet-of-Things on the battlefield; "cognitive jamming" in electronic warfare (e.g., the Army Rapid Capabilities Office recently issued a Request for Information (RFI) for AI and machine learning algorithms in support of EW); and cyber defensive and offensive "reasoning systems." While not even the winner of DARPA's recent Cyber Challenge 2016 demonstrated anything close to *AlphaGo*-level performance,[63] it can be argued that this first-of-its-kind challenge is analogous to DARPA's 2004 "Driverless Car" challenge.[64] For *that* first-of-its-kind challenge, not one of the contestants was able to finish more than 5 percent of the challenge course. Yet, a dozen or so years later, self-driving cars are now commercially available. There is no reason to believe that cyber-AI will not follow a similar trajectory.

[63] http://archive.darpa.mil/cybergrandchallenge/.

[64] J. Hooper, "From Darpa Grand Challenge 2004: DARPA's Debacle in the Desert," *Popular Science*, 4 June 2004. http://www.popsci.com/scitech/article/2004-06/darpa-grand-challenge-2004darpas-debacle-desert.

CNA

DIS-2017-U-016388-Final-2

CNA is a not-for-profit research organization
that serves the public interest by providing
in-depth analysis and result-oriented solutions
to help government leaders choose
the best course of action
in setting policy and managing operations.

Nobody gets closer—
to the people, to the data, to the problem.

CNA
ANALYSIS & SOLUTIONS

www.cna.org • 703-824-2000

3003 Washington Boulevard, Arlington, VA 22201